

OIPE

RECEIVED
FEB 06 2002
TECH CENTER 1600/2900

DATE: 11/27/2001
TIME: 11:20:23

ENTERED

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/863,475A

Input Set : A:\09863475A.raw.txt
Output Set: N:\CRF3\11212001\I863475A.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:

- 6 (i) APPLICANT: LOWE, JOHN B.
8 (ii) TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
9 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
10 GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
11 OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE

STRUCTURES

- 13 (iii) NUMBER OF SEQUENCES: 14
15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
17 P.C.

18 (B) STREET: 1755 Jefferson Davis Highway, Fourth Floor
19 (C) CITY: Arlington

20 (D) STATE: Virginia
21 (E) COUNTRY: U.S.A.

22 (F) ZIP: 22202

- 24 (v) COMPUTER READABLE FORM:

25 (A) MEDIUM TYPE: Floppy disk
26 (B) COMPUTER: IBM PC compatible
27 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
28 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

- 30 (vi) CURRENT APPLICATION DATA:

C--> 31 (A) APPLICATION NUMBER: US/09/863,475A
C--> 32 (B) FILING DATE: 24-May-2001
33 (C) CLASSIFICATION:

- 35 (vii) PRIOR APPLICATION DATA:

36 (A) APPLICATION NUMBER: 07/914,281
37 (B) FILING DATE: 20-JUL-1992

- 39 (viii) ATTORNEY/AGENT INFORMATION:

40 (A) NAME: Lavalleye, Jean-Paul M. P.
41 (B) REGISTRATION NUMBER: 31,451
42 (C) REFERENCE/DOCKET NUMBER: 2363-060-55

- 44 (ix) TELECOMMUNICATION INFORMATION:

45 (A) TELEPHONE: (703)521-4500
46 (B) TELEFAX: (703)486-2347
47 (C) TELEX: 248855 OPAT UR

- 50 (2) INFORMATION FOR SEQ ID NO: 1:

- 52 (i) SEQUENCE CHARACTERISTICS:

53 (A) LENGTH: 2043 base pairs
54 (B) TYPE: nucleic acid
55 (C) STRANDEDNESS: unknown
56 (D) TOPOLOGY: unknown

- 58 (ii) MOLECULE TYPE: cDNA

- 60 (iv) ANTI-SENSE: NO

- 64 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

66 AGGAAACCTG CCATGGCCTC CTGGTGAGCT GTCCTCATCC ACTGCTCGCT GCCTCTCCAG 60
68 ATACTCTGAC CCATGGATCC CCTGGGTGCA GCCAAGCCAC AATGGCCATG GCGCCGCTGT 120

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/863,475A

DATE: 11/27/2001
TIME: 11:20:23

Input Set: A:\09863475A.raw.txt
Output Set: N:\CRF3\11212001\I863475A.raw

```

70 CTGGCCGCAC TGCTATTTC A GCTGCTGGTG GCTGTGTGTT TCTTCTCCTA CCTGCGTGTG 180
72 TCCCCGAGACG ATGCCACTGG ATCCCCTAGG GCTCCCAGTG GGTCTCCCG ACAGGACACC 240
74 ACTCCCACCC GCCCACCCTT CCTGATCCTG CTATGGACAT GGCCTTTCCA CATCCCTGTG 300
76 GCTCTGTCCC GCTGTTTACA GATGGTGCCC GGCACAGCCG ACTGCCACAT CACTGCCGAC 360
78 CGCAAGGTGT ACCCAGAGG AGACACGGTC ATCGTGACC ACTGGGATAT CATGTCCAAC 420
80 CCTAAGTCAC GCCTCCACC TTCCCCGAGG CCGCAGGGG AGCGCTGGAT CTGGTTCAAC 480
82 TTGGAGCCAC CCCCTAACTG CCAGCACCTG GAAGCCCTGG ACAGATACTT CAATCTCACC 540
84 ATGTCTTACC GCAGCGACTC CGACATCTTC ACGCCCTACG GCTGGCTGGA GCCGTGGTCC 600
86 GGCCAGCCTG CCCACCACC GCTCAACCTC TCGGCCAAGA CCGAGCTGGT GGCCTGGGCG 660
88 GTGTCCAACT GGAAGCCGGA CTCAGCCAGG GTGCGCTACT ACCAGAGCCT GCAGGCTCAT 720
90 CTCAAGGTGG ACGTGTACGG ACGCTCCAC AAGCCCTGCG CCAAGGGGAC CATGATGGAG 780
92 ACGCTGTCCC GGTACAAGTT CTACCTGGCC TTCGAGAACT CCTTGCAACC CGACTACATC 840
94 ACCGAGAAGC TGTGGAGGAA CGCCCTGGAG GCCTGGGCGG TGCCCGTGGT GCTGGGCCCC 900
96 AGCAGAAGCA ACTACGAGAG GTTCTGCCA CCCGACGCT TCATCCACGT GGACGACTTC 960
98 CAGAGCCCCA AGGACCTGGC CCGGTACCTG CAGGAGCTGG ACAAGGACCA CGCCCGCTAC 1020
100 CTGAGCTACT TTCGCTGGCG GGAGACGCTG CGGCCTCGCT CTTTACGCTG GGCCTGGAT 1080
102 TTCTGCAAGG CCTGCTGGAA ACTGCAGCAG GAATCCAGGT ACCAGACGGT GCGCAGCATA 1140
104 GCGGCTTGGT TCACCTGAGA GGCCGCGATG GTGCCTGGGC TGCCGGGAAC CTCATCTGCC 1200
106 TGGGGCTTCA CCTGCTGGAG TCCTTTGTGG CCAACCCTCT CTCTTACCTG GGACCTCACA 1260
108 CGCTGGGCTT CACGGCTGCC AGGAGCCTCT CCCCTCCAGA AGACTTGCCT GCTAGGGACC 1320
110 TCGCCTGCTG GGGACCTCGC CTGTTGGGGA CCTCACCTGC TGGGGACCTC ACCTGCTGGG 1380
112 GACCTTGGCT GCTGGAGGCT GCACCTACTG AGGATGTGCG CGGTGCGGGA CTTTACCTGC 1440
114 TGGGACCTGC TCCAGAGAC CTTGCCACAC TGAATCTCAC CTGCTGGGGA CCTCACCTG 1500
116 GAGGGCCCTG GGCCCTGGGG AACTGGCTTA CTTGGGGCCC CACCCGGGAG TGATGGTTCT 1560
118 GGCTGATTTG TTTGTGATGT TGTAAGCCGC CTGTGAGGGG TGCAGAGAGA TCATCACGGC 1620
120 ACGGTTTCCA GATGTAATAC TGCAAGGAAA AATGATGACG TCAGGGGTTA AGAGCTCAGA 1680
122 GTTGGTCCCA TGGGTAAAGA GCTCACCACA GGTCTCTACC TTAGGGTACAG GTGGGATCGA 1740
124 GTTCAGACAG GTCCAAGTTC AAGCCCAGGA CCACCCTTA TAGGGTACAG GTGGGATCGA 1800
126 CTGTAAATGA GGAATTCTGG AACATTCCAA ATATTCTGGG GTTGAGGGAA ATTGCTGCTG 1860
128 TCTACAAAAT GCCAAGGGTG GACAGGCGCT GTGGCTCACG CCTGTAATTC CAGCACTTTG 1920
130 GGAGGCTGAG GTAGGAGGAT TGATTGAGGC CAAGAGTTAA AGACCAGCCT GGTCAATATA 1980
132 GCAAGACCAC GTCTCTAAAT AAAAAATAAT AGGCCGGCCA GGAAAAA AAAA AAAAAA 2040
134 AAA

```

136 (2) INFORMATION FOR SEQ ID NO: 2:

138 (i) SEQUENCE CHARACTERISTICS:

139 (A) LENGTH: 361 amino acids

140 (B) TYPE: amino acid

141 (D) TOPOLOGY: unknown

143 (ii) MOLECULE TYPE: protein

147 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

149 Met Asp Pro Leu Gly Ala Ala Lys Pro Gln Trp Pro Trp Arg Arg Cys
150 1 5 10 15
152 Leu Ala Ala Leu Leu Phe Gln Leu Leu Val Ala Val Cys Phe Phe Ser
153 20 25 30
155 Tyr Leu Arg Val Ser Arg Asp Asp Ala Thr Gly Ser Pro Arg Ala Pro
156 35 40 45
158 Ser Gly Ser Ser Arg Gln Asp Thr Thr Pro Thr Arg Pro Thr Leu Leu
159 50 55 60
161 Ile Leu Leu Trp Thr Trp Pro Phe His Ile Pro Val Ala Leu Ser Arg

```

11/27/01

RAW SEQUENCE LISTING

DATE: 11/27/2001

PATENT APPLICATION: US/09/863,475A

TIME: 11:20:23

Input Set: A:\09863475A.raw.txt

Output Set: N:\CRF3\11212001\I863475A.raw

```

162      65      70      75      80
164      Cys Ser Glu Met Val Pro Gly Thr Ala Asp Cys His Ile Thr Ala Asp
165      85      90      95
167      Arg Lys Val Tyr Pro Gln Ala Asp Thr Val Ile Val His His Trp Asp
168      100      105      110
170      Ile Met Ser Asn Pro Lys Ser Arg Leu Pro Pro Ser Pro Arg Pro Gln
171      115      120      125
173      Gly Gln Arg Trp Ile Trp Phe Asn Leu Glu Pro Pro Asn Cys Gln
174      130      135      140
176      His Leu Glu Ala Leu Asp Arg Tyr Phe Asn Leu Thr Met Ser Tyr Arg
177      145      150      155      160
179      Ser Asp Ser Asp Ile Phe Thr Pro Tyr Gly Trp Leu Glu Pro Trp Ser
180      165      170      175
182      Gly Gln Pro Ala His Pro Pro Leu Asn Leu Ser Ala Lys Thr Glu Leu
183      180      185      190
185      Val Ala Trp Ala Val Ser Asn Trp Lys Pro Asp Ser Ala Arg Val Arg
186      195      200      205
188      Tyr Tyr Gln Ser Leu Gln Ala His Leu Lys Val Asp Val Tyr Gly Arg
189      210      215      220
191      Ser His Lys Pro Leu Pro Lys Gly Thr Met Met Glu Thr Leu Ser Arg
192      225      230      235      240
194      Tyr Lys Phe Tyr Leu Ala Phe Glu Asn Ser Leu His Pro Asp Tyr Ile
195      245      250      255
197      Thr Glu Lys Leu Trp Arg Asn Ala Leu Glu Ala Trp Ala Val Pro Val
198      260      265      270
200      Val Leu Gly Pro Ser Arg Ser Asn Tyr Glu Arg Phe Leu Pro Pro Asp
201      275      280      285
203      Ala Phe Ile His Val Asp Asp Phe Gln Ser Pro Lys Asp Leu Ala Arg
204      290      295      300
206      Tyr Leu Gln Glu Leu Asp Lys Asp His Ala Arg Tyr Leu Ser Tyr Phe
207      305      310      315      320
209      Arg Trp Arg Glu Thr Leu Arg Pro Arg Ser Phe Ser Trp Ala Leu Asp
210      325      330      335
212      Phe Cys Lys Ala Cys Trp Lys Leu Gln Gln Glu Ser Arg Tyr Gln Thr
213      340      345      350
215      Val Arg Ser Ile Ala Ala Trp Phe Thr
216      355      360
218 (2) INFORMATION FOR SEQ ID NO: 3:
220      (i) SEQUENCE CHARACTERISTICS:
221          (A) LENGTH: 1500 base pairs
222          (B) TYPE: nucleic acid
223          (C) STRANDEDNESS: unknown
224          (D) TOPOLOGY: unknown
226      (ii) MOLECULE TYPE: cDNA
228      (iv) ANTI-SENSE: NO
232      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
234 CCTTCCCTTG TAGACTCTTC TTGGAATGAG AAGTACCGAT TCTGCTGAAG ACCTCGCGCT      60
236 CTCAGGCTCT, GGGAGTTGGA ACCCTGTACC TTCCTTTCCT CTGCTGAGCC CTGCCTCCTT      120
238 AGGCAGGCCA GAGCTCGACA GAACTCGGTT GCTTTGCTGT TTGCTTTGGA GGGAACACAG      180

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/863,475A

DATE: 11/27/2001

TIME: 11:20:23

Input Set: A:\09863475A.raw.txt

Output Set: N:\CRF3\11212001\I863475A.raw

```

240 CTGACGATGA GGCTGACTTT GAACTCAAGA GATCTGCTTA CCCCAGTCTC CTGGAATTAA 240
242 AGGCCGTGAC TACATTTGCC TGGACCTAAG ATTTTCATGA TCACTATGCT TCAAGATCTC 300
244 CATGTCAACA AGATCTCCAT GTCAAGATCC AAGTCAGAAA CAAGTCTTCC ATCCTCAAGA 360
246 TCTGGATCAC AGGAGAAAAT AATGAATGTC AAGGGAAAAG TAATCCTGTT GATGCTGATT 420
248 GTCTCAACCG TGGTTGTCGT GTTTTGGGAA TATGTCAACA GAATTCCAGA GGTTGGTGAG 480
250 AACAGATGGC AGAAGGACTG GTGGTTCCCA AGCTGGTTTA AAAATGGGAC CCACAGTTAT 540
252 CAAGAAGACA ACGTAGAAGG ACGGAGAGAA AAGGGTAGAA ATGGAGATCG CATTGAAGAG 600
254 CCTCAGCTAT GGGACTGGTT CAATCCAAAG AACC GCCCGG ATGTTTTGAC AGTGACCCCG 660
256 TGGAAGGCGC CGATTGTGTG GGAAGGCACT TATGACACAG CTCTGCTGGA AAAGTACTAC 720
258 GCCACACAGA AACTCACTGT GGGGCTGACA GTGTTTGCTG TGGGAAAGTA CATTGAGCAT 780
260 TACTTAGAAG ACTTTCTGGA GTCTGCTGAC ATGTACTTCA TGTTTGCCA TCGGGTCATA 840
262 TTTTACGTCA TGATAGACGA CACCTCCCGG ATGCTGTGCG TGCACCTGAA CCCTCTACAT 900
264 TCCTTACAAG TCTTTGAGAT CAGGTCTGAG AAGAGGTGGC AGGATATCGA CATGATGCGC 960
266 ATGAAGACCA TTGGGGAGCA CATCCTGGCC CACATCCAGC ACGAGGTCTG CTTCTCTTTC 1020
268 TGCATGGACG TGGATCAAGT CTTTCAAGAC AACTTCGGGG TGGAAACTCT GGGCCAGCTG 1080
270 GTAGCACAGC TCCAGGCCTG GTGGTACAAG GCCAGTCCCG AGAAGTTCAC CTATGAGAGG 1140
272 CGGGAACGTG CGGCCGCGTA CATTCCATTC GGAGAGGGGG ATTTTACTA CCACGCGGCC 1200
274 ATTTTGGAG GAACGCCTAC TCACATTCTC AACCTACCA GGGAGTGCTT TAAGGGGATC 1260
276 CTCCAGGACA AGAAACATGA CATAGAAGCC CAGTGGCATG ATGAGAGCCA CCTCAACAAA 1320
278 TACTTCCTTT TCAACAAACC CACTAAAATC CTATCTCCAG AGTATTGCTG GGACTATCAG 1380
280 ATAGGCCTGC CTTTCAATAT TAAAAGTGTC AAGGTAGCTT GGCAGACAAA AGAGTATAAT 1440
282 TTGGTTAGAA ATAATGTCTG ACTTCAAATT GTGATGAAA CTTGACACTA TTTCTAACCA 1500
285 (2) INFORMATION FOR SEQ ID NO: 4:
287 (i) SEQUENCE CHARACTERISTICS:
288 (A) LENGTH: 394 amino acids
289 (B) TYPE: amino acid
290 (D) TOPOLOGY: unknown
292 (ii) MOLECULE TYPE: protein
296 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
298 Met Ile Thr Met Leu Gln Asp Leu His Val Asn Lys Ile Ser Met Ser
299 1 5 10 15
301 Arg Ser Lys Ser Glu Thr Ser Leu Pro Ser Ser Arg Ser Gly Ser Gln
302 20 25 30
304 Glu Lys Ile Met Asn Val Lys Gly Lys Val Ile Leu Leu Met Leu Ile
305 35 40 45
307 Val Ser Thr Val Val Val Val Phe Trp Glu Tyr Val Asn Arg Ile Pro
308 50 55 60
310 Glu Val Gly Glu Asn Arg Trp Gln Lys Asp Trp Trp Phe Pro Ser Trp
311 65 70 75
313 Phe Lys Asn Gly Thr His Ser Tyr Gln Glu Asp Asn Val Glu Gly Arg
314 85 90 95
316 Arg Glu Lys Gly Arg Asn Gly Asp Arg Ile Glu Glu Pro Gln Leu Trp
317 100 105 110
319 Asp Trp Phe Asn Pro Lys Asn Arg Pro Asp Val Leu Thr Val Thr Pro
320 115 120 125
322 Trp Lys Ala Pro Ile Val Trp Glu Gly Thr Tyr Asp Thr Ala Leu Leu
323 130 135 140
325 Glu Lys Tyr Tyr Ala Thr Gln Lys Leu Thr Val Gly Leu Thr Val Phe
326 145 150 155 160

```

11/27/01

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/863,475A

DATE: 11/27/2001
TIME: 11:20:23

Input Set: A:\09863475A.raw.txt
Output Set: N:\CRF3\11212001\I863475A.raw

328 Ala Val Gly Lys Tyr Ile Glu His Tyr Leu Glu Asp Phe Leu Glu Ser 175
329 165 170
331 Ala Asp Met Tyr Phe Met Val Gly His Arg Val Ile Phe Tyr Val Met 190
332 180 185
334 Ile Asp Asp Thr Ser Arg Met Pro Val Val His Leu Asn Pro Leu His 205
335 195 200
337 Ser Leu Gln Val Phe Glu Ile Arg Ser Glu Lys Arg Trp Gln Asp Ile 220
338 210 215
340 Ser Met Met Arg Met Lys Thr Ile Gly Glu His Ile Leu Ala His Ile 240
341 225 230 235
343 Gln His Glu Val Asp Phe Leu Phe Cys Met Asp Val Asp Gln Val Phe 255
344 245 250
346 Gln Asp Asn Phe Gly Val Glu Thr Leu Gly Gln Leu Val Ala Gln Leu 270
347 260 265
349 Gln Ala Trp Trp Tyr Lys Ala Ser Pro Glu Lys Phe Thr Tyr Glu Arg 285
350 275 280
352 Arg Glu Leu Ser Ala Ala Tyr Ile Pro Phe Gly Glu Gly Asp Phe Tyr 300
353 290 295
355 Tyr His Ala Ala Ile Phe Gly Gly Thr Pro Thr His Ile Leu Asn Leu 320
356 305 310 315
358 Thr Arg Glu Cys Phe Lys Gly Ile Leu Gln Asp Lys Lys His Asp Ile 335
359 325 330
361 Glu Ala Gln Trp His Asp Glu Ser His Leu Asn Lys Tyr Phe Leu Phe 350
362 340 345
364 Asn Lys Pro Thr Lys Ile Leu Ser Pro Glu Tyr Cys Trp Asp Tyr Gln 365
365 355 360
367 Ile Gly Leu Pro Ser Asp Ile Lys Ser Val Lys Val Ala Trp Gln Thr 380
368 370 375
370 Lys Glu Tyr Asn Leu Val Arg Asn Asn Val 390
371 385 390

373 (2) INFORMATION FOR SEQ ID NO: 5:

375 (i) SEQUENCE CHARACTERISTICS:

- 376 (A) LENGTH: 8174 base pairs
377 (B) TYPE: nucleic acid
378 (C) STRANDEDNESS: unknown
379 (D) TOPOLOGY: unknown

381 (ii) MOLECULE TYPE: DNA (genomic)

383 (iv) ANTI-SENSE: NO

387 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

389	GAATTCATC	GTGGCAAGGG	CAGCCTGAAT	GGATGATGTA	ACCTGGGGTC	CTTTCAATGG	60
391	AGGGCCAGAC	TCCTGGGTCT	AGGGGATGAG	GGAGGGGAGG	ATCGGGTTAG	CTGGGACCCA	120
393	GGTGAAAGGG	GCTGGGGGCC	CACATTCCTG	AGTCTCAGAG	AGAAGGATCT	GGGGTCTCAA	180
395	GCACCTGAGT	CGGAGGGAGG	AGGGGTGCTG	GGCTCCTGGA	AAAACCACCT	CTTGGACCAT	240
397	CTATGCAGAT	CACGCAGAAC	AAGAGAAATT	TCTGCGCCCC	ATCTGAATTT	CTAAGTTTGG	300
399	GGGGAGGGCG	TGATCTGACA	CTGAGGTTC	TTGATCCTCA	GCAAGGCGGC	AATTGCTGTA	360
401	TGAAAGAAGC	GACCGCATCT	GAGACACAAG	TATCCTGCCT	TGGAAGCCTC	TCACCTGGCC	420
403	GTGGGCCAAC	CTCAACCTCA	TCTGTCCCTG	CTCAGATGCT	CAGACCCTGG	ACATCCCAGC	480
405	CTCCTCCTCC	CTGATGCAAT	CCTGGTGTTC	CTTTCACCAG	AGAAGCCATC	CCAGGCCAGC	540
407	GCAGGTGCTC	CTGAAATAAC	CTGGGGGGAG	GGGTGGCTGA	AAGTCCCTGA	CTGGAGTTGG	600

11/27/01

VERIFICATION SUMMARY
PATENT APPLICATION: US/09/863,475A

DATE: 11/27/2001
TIME: 11:20:24

Input Set: A:\09863475A.raw.txt
Output Set: N:\CRF3\11212001\I863475A.raw

L:31 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:32 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]

11/27/01